DOCUMENT RESUME

ED 315 054 IR 014 168

TITLE	Input and Output Mechanisms and Devices. Phase I: Adding Voice Output to a Speaker-Independent Recognition System.
INSTITUTION	Scott Instruments Corp., Denton, TX.
SPONS AGENCY	Department of Education, Washington, DC.
REPORT NO	RFP-NIE-R-85-0006
PUB DATE	Mar 85
CONTRACT	400-85-1021
NOTE	7p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Authoring Aids (Programing); *Courseware; *Input Output Devices; *Microcomputers; Programing Languages; Speech Compression; *Speech Synthesizers
IDENTIFIERS	*Apple Macintosh; PASCAL Programing Language

ABSTRACT

This project was designed to develop techniques for adding low-cost speech synthesis to educational software. Four tasks were identified for the study: (1) select a microcomputer with a built-in analog-to-digital converter that is currently being used in educational environments; (2) determine the feasibility of implementing expansion and playback routines of compressed speech files on the targeted microcomputer; (3) test the target microcomputer using various bit rates; and (4) determine the feasibility of incorporating speech output capability into an authoring system currently being developed for speech recognition. The Apple Macintosh was selected, and routines were developed in the Pascal programming language for expanding and playing speech files at various data rates. It was found that the Macintosh is capable of producing a range of synthesized speech and that it is necessary to provide the courseware developer with a selection of bit rates. (MES)

5/11/30

ED315054

Department of Education, RFP NIE-R-85-0006

Topic 9: Input and Output Mechanisms and Devices

Phase I - Adding Voice Output to a Speaker-Independent Recognition System

Scott Instruments Corporation
1111 Willow Springs Drive
Denton, Texas 76205

This material is based upon work supported by the U.S. Department of Education under contract number 400-85-1021. Any opinions, finding, conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views or policies of the Department of Education.

IR014168

Scott Instruments Corporation
1111 Willow Springs Drive
Denton, Texas 76205

Dr. Gary Goodman, Director of Research

Department of Education, RFP NIE-R-85-0006

Topic 9: Input and Output Mechanisms and Devices

Phase I - Adding Voice Output to a Speaker-Independent Recognition System

Final Report

Abstract. The major goal of the original Phase I proposal was to develop a low-cost means for adding speech output to educational software. The planned vehicle for allowing that was a low-cost speech recognition card for microcomputers commonly found in educational environments. Since funding for the recognition card was terminated (ED did not fund the Phase II proposal), it was decided that development of the synthesis capability should proceed on a particular educational machine. We chose the Apple Macintosh computer. This decision was made for two reasons: (1) the Macintosh computer has a built-in A/D convertor capable of playing high quality speech at no additional cost to the user, and (2) Scott Instruments is in the process of developing an authoring system on the Macintosh computer which could provide an easy-to-use method for integrating speech output into educational software.

It has been shown that the Macintosh computer is capable of producing a range of resynthesized speech, from pleasing the most discriminating of ears (at 32 kbs.), to meeting the needs of those more concerned with quantity of speech (22 minutes of speech per disk at 4.8 kbs.). More importantly, the study has indicated that it is probably necessary to provide the courseware developer with a selection of bit rates, all of which should be available even within a particular program. Instructions and prompts should be generated at low bit rates for efficiency and articulation models should be generated at high bit rates for accuracy.

Key Words: Speech output device, CAI, Interactive learning methods

Anticipated Results: The above findings have led to a Phase II proposal for developing an authoring system capable of integrating speech output into courseware developed for the Macintosh computer. The proposed system will require additional hardware for the developer to enable him to digitize and compress speech files at a variety of different bit rates, but will require no additional hardware to run the courseware. A model of the authoring system capable of being modified to fit the requirements is currently under development and the extent of the modifications will be addressed in the Phase II proposal.

Final Report. The current Phase I SBIR project was designed to develop techniques for adding low-cost speech synthesis to educational software. The original Phase I proposal was intended to fund the modification of a low-cost speech recognition module to support speech synthesis as well as speech recognition. The design of the recognition module was supported by ED SBIR grant number 300-84-0174 entitled, "Phase I - Feature based Recognition for Speaker-Independent Voice Control of Microcomputers." A Phase II proposal was submitted as a result of this study in March of 1985. The Phase II proposal was rejected by the Department of Education, but the current Phase I feasibility study was accepted. Since the support for the development of the recognition hardware was terminated, an alternative approach toward the development of a low-cost speech synthesis capability for educational software was adopted. Instead of developing general purpose hardware for recognition and synthesis, it was determined that the most cost-effective way to provide low-cost speech synthesis to the educational market was to develop a synthesis capability for a microcomputer with a built-in digital-to-analog convertor capable of producing intelligible speech over a broad range of speech compression rates.

The effort supported by this Phase I grant has resulted in a demonstration of the feasibility of producing highly intelligible speech output within a simple-to-use authoring system framework which is capable of generating courseware, and which requires no additional hardware for student stations. This report will discuss the progress made toward the goal of enabling educational courseware developers, as well as non-computer oriented teachers, to develop useful classroom courseware rapidly and efficiently with speech output capabilities.

Identification and Significance of the Problem

Speech input and output may be the most powerful adjunct technologies that can be added to computer assisted instructional programs. Some of the advantages of speech technology in this context are obvious, others more subtle. One obvious and well-known example of the power of speech technology is Texas Instruments' popular and highly successful Speak and Spell. This represents a perfect marriage between technology and education. We believe that there is no more effective way to teach spelling than through speech. Other obvious examples of appropriate educational applications for speech technology include: the teaching of foreign languages, teaching English as a second language, the training of speech pathologists and audiologists, and the rehabilitation of persons with communication disorders. Speech technology is capable of involving the student with the instructional machine in a way that cannot otherwise be obtained.

Given that speech technology can play an important and unique role in Computer Assisted Instruction (CAI), the problem is one of getting the technology into the hands of the educator. One way of accomplishing this is to provide the educator with complete, ready-to-run, off-the-shelf software. This represents the simplest option for the educator. However, the technical know-how required to incorporate speech technology into educational software, as well as the additional hardware costs associated with adding a speech output capability to each student's workstation computer begs the following questions: 1. How does an institution justify the purchase of a speech module for classroom use if no educational software exists? 2. Why would educational software developers incorporate speech technology into their offering if no one has the appropriate hardware for voice output? 3. How do we deliver the educational power of speech technology into the hands of the educator in a way that allows for the creative use of the technology as a dynamic tool in the educational process?

The answer is apparently not in the development of off-the-shelf, static lessons which cannot be modified by the educator. Courseware purchased over a period of time from many different developers will obviously lack continuity. The funding cycles of educational institutions ensure that educators will never be current in their courseware libraries. In fact, the lack of courseware can render an expensive computer laboratory useless to all but a few.

The challenge is, therefore, to develop a means through which any educator may develop courseware with a speech input or output capability. Such a system must require minimum training: The goal is to provide a tool for all educators, not just those with computer skills. More important, the system must be cost-effective in its production of educational software. Additionally, the system must be designed to take advantage of future enhancements in the student workstation without requiring such improvements in the present. In other words, both the courseware development system and the student workstation must not be allowed to make each other obsolete.

What has been described is an authoring system capable of handling the integration of speech input/output into lessons. The use of the courseware should require as little additional capital expense as possible in order for the system to become widely utilized. The present Phase I feasibility study has been directed at investigating the integration of speech technology into an authoring system which is simple enough for the inexperienced computer users to use effectively.

Scott Instruments has been developing voice-based authoring systems since 1981. Our first system was designated the VBLS®, voice-based learning system. This system utilized a $900 speech recognition system available for the Apple II series computers. The most successful application of the VBLS system has been in foreign languages training. Systems are currently being used daily across the state of Oklahoma for teaching German by satellite, a program established by Dr. Harry Wohlert, a professor in the Foreign Languages department at Oklahoma State University. The same system was proven effective in the development of rehabilitative courseware for speech and hearing disorders during a Phase I study supported by the National Institute of Neurological and Communicative Disorders, grant number 1 R43 NS 21417-01 CMS (OD), awarded September of 1984. These studies, and the experiences associated with having approximately two hundred VBLS systems in the field, has taught us a great deal about the use of speech technology in education.

The most important items are: (1) the cost per student station must be as low as possible, (2) the system must be simple enough for anyone to use, (3) high resolution graphics is often a requirement for many applications, and (4) the system should have speech synthesis as well as speech recognition capabilities.

Phase I - Goals and accomplishments

As mentioned above, because the current study was, in part, based on the funding of a companion study (which was not funded), the feasibility study was broken into the following tasks:

(1) select a target microcomputer currently being used in educational environments, with a built in analog-to-digital convertor;

(2) determine the feasibility of implementing expansion and playback routines of compressed speech files on the target microcomputer;

®VBLS is a registered trademark of Scott Instruments Corporation.

(3) test target microcomputer using various bit rates;
(4) determine feasibility of incorporating speech output capability into an authoring system currently being developed for speech recognition.

Target Microcomputer

The microcomputer selected was the Apple Macintosh. The Macintosh computer has a built in A/D convertor, is becoming a popular computer in educational environments, uses the powerful Motorola 68000 microprocessor, has high-resolution graphics capabilities, and is a machine on which we are currently developing an authoring system which can be adapted to utilize speech synthesis in a simple and effective way.

Speech synthesis on the Macintosh

Routines were developed in the Pascal programming language for expanding and playing speech files previously compressed on a VAX 11/750 development machine located at Scott Instruments. The data rate for the compressed files was 16 kilobits (kbs.). The quality of the compressed speech can be heard on the enclosed demonstration tape.

The 16-kbs. per second data rate represents the simplest of the expansion routines. The total amount of speech that can be stored on a Macintosh microfloppy at the 16-kbs. data rate is seven minutes. However, since we developed non-real time compression and expansion models on our development system that operated at data rates as low as 2.4 kbs., the decision was made to attempt to push data rates as low as possible on the Macintosh computer. Thus far, assembly language routines have been developed and debugged using our own 68000-based development system for data rates as low as 4.8 kbs. per second.

The following table shows the data rates for which 68000 assembly code has been developed and tested along with the number of minutes of continuous speech that can be stored on a single Macintosh microfloppy disk.

Data Rate	Number of minutes per disk
32 kbs	3.3 min
16 kbs	7 min
9.6 kbs	11 min
4.8 kbs	22 min

These assembly language routines can be quickly adapted to the Macintosh computer for all of the above listed speech compression rates. The enclosed demonstration tape contains examples of the speech quality of each data rate. The sentences recorded were taken from the Arizona Articulation Proficiency Scale Sentence Test. The twenty-five sentences from the Arizona test were recorded

at each of the four bit rates for a total of 100 recordings. Preliminary analysis of the speech quality was done by having five speech pathologists from the North Texas State University Division of Communication Disorders listen to the tapes and judge whether the target phones designated in the test were adequate as models for use in articulation therapy. The consensus was that the 32-kbs. model was completely adequate, and the 16- and 9.6-kbs. models were marginal. The 4.8-kbs. rate was judged inappropriate for use as speech models for teaching proper articulation, but was adequate for verbal instructions to a student.

One important finding in the pilot study was that the persons evaluating the synthesis objected more to unnatural sounding transients in the recorded speech than to the speech quality and intelligibilty. This indicates the need for developing speech editing tools for the developer that would enable him to modify parameters such as amplitude envelope, the endpoints of the desired phrase, and possibly even the pitch contour. These editing tools are all possible within the constraints of the authoring system currently under development.

Summary and Conclusions

The major goal of the original Phase I proposal was to develop a low-cost means for adding speech output to educational software. The planned vehicle for allowing that was a low-cost speech recognition card for microcomputers commonly found in educational environments. Since funding for the recognition card was terminated, the project was altered to develop synthesis capability on a particular educational machine, the Apple Macintosh computer. This decision was made for two reasons: (1) the Macintosh computer has a built-in A/D convertor capable of playing high quality speech at no additional cost to the user, and (2) Scott Instruments is in the process of developing an authoring system on the Macintosh computer which could provide an easy-to-use method for integrating speech output into educational software.

In order to determine whether or not the Macintosh computer could generate high-quality speech synthesis in real-time, expansion routines were developed in assembly language on a 68000-based development machine at Scott Instruments. Since the required quality of speech was unknown, algorithms capable of generating speech compressed at 32 kbs., 16 kbs., 9.6 kbs., and 4.8 kbs. were all developed and tested. It was found in an informal test using trained speech pathologists as judges, that higher bit rates would be required for some applications and lower bit rates would be adequate for others.

In conclusion, we have shown that the Macintosh computer is capable of producing a range of resynthesized speech, from pleasing the most discriminating of ears (at 32 kbs.), to meeting the needs of those more concerned with quantity of speech (22 minutes of speech per disk at 4.8 kbs). More importantly, the study has indicated that it is probably necessary to provide the courseware developer with a selection of bit rates, all of which should be available even within a particular program. Instructions and prompts should be generated at low bit rates for efficiency and articulation models should be generated at high bit rates for accuracy.

The above findings have led to a Phase II proposal for developing an authoring system capable of integrating speech output into courseware developed for the Macintosh computer. The proposed system will require additional hardware for the developer to enable him to digitize and compress speech files at a variety of different bit rates, but will require no additional hardware to run the courseware. A model of the authoring system capable of being modified to fit the requirements is currently under development and the extent of the modifications will be addressed in the Phase II proposal.